

Regulatory genes in the ancestral chordate genomes

Yutaka Satou · Shuichi Wada · Yasunori Sasakura ·
Nori Satoh

Received: 29 October 2007 / Accepted: 10 April 2008 / Published online: 5 September 2008
© The Author(s) 2008

Abstract Changes or innovations in gene regulatory networks for the developmental program in the ancestral chordate genome appear to be a major component in the evolutionary process in which tadpole-type larvae, a unique characteristic of chordates, arose. These alterations may include new genetic interactions as well as the acquisition of new regulatory genes. Previous analyses of the *Ciona* genome revealed that many genes may have emerged after the divergence of the tunicate and vertebrate lineages. In this paper, we examined this possibility by examining a second non-vertebrate chordate genome. We conclude from this analysis that the ancient chordate included almost the same repertoire of regulatory genes, but less redundancy than extant vertebrates, and that approximately 10% of vertebrate regulatory genes were innovated after the

emergence of vertebrates. Thus, refined regulatory networks arose during vertebrate evolution mainly as preexisting regulatory genes multiplied rather than by generating new regulatory genes. The inferred regulatory gene sets of the ancestral chordate would be an important foundation for understanding how tadpole-type larvae, a unique characteristic of chordates, evolved.

Keywords *Branchiostoma floridae* · *Ciona intestinalis* · Evolution · Regulatory genes

Introduction

Emergence of the tadpole-like body plan in an ancestral chordate is still a big mystery in the evolutionary biology. To know what happened in the genome of the ancestral chordate for acquiring the chordate body plan, regulatory genes in the tunicate genome that encode transcription factors and signaling ligands have been comprehensively identified (Hino et al. 2003; Satou et al. 2003a, b; Wada et al. 2003; Yagi et al. 2003; Yamada et al. 2003) because the tunicates and vertebrates are a sister group within chordates (Delsuc et al. 2006). The regulatory gene set identified in these previous studies may comprise that possessed by an ancestral chordate, most part of which is shared with vertebrates. Vertebrate genomes contain additional regulatory genes not found in the tunicate genome, such as *Tbx4/5*, *Gbx*, and *wnt11* (Dehal et al. 2002). Some are a result of gene loss in the tunicate genome, as protostomes and vertebrates share clear orthologs (e.g., *Wnt4*). Others appear to have rapidly evolved in *Ciona*, making their phylogenetic positions obscure. The remaining may be genes that were innovated in the vertebrate lineage. Such classification is apparently essential for

Communicated by M. Hammerschmidt.

Yutaka Satou, Shuichi Wada and Yasunori Sasakura contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00427-008-0219-y) contains supplementary material, which is available to authorized users.

Y. Satou (✉) · N. Satoh
Department of Zoology, Graduate School of Science,
Kyoto University,
Sakyo, Kyoto 606-8501, Japan
e-mail: yutaka@ascidian.zool.kyoto-u.ac.jp

S. Wada
Department of Bioscience, Faculty of Bioscience,
Nagahama Institute of Bio-Science and Technology,
Nagahama, Shiga 526-0829, Japan

Y. Sasakura
Shimoda Marine Research Center, University of Tsukuba,
5-10-1,
Shimoda, Shizuoka 415-0025, Japan

understanding of the origin of chordates. Recent decoding of the amphioxus genome gave us an opportunity to solve this problem because the cephalochordates are thought to be the most basal group within chordates. In this study, we compare the major regulatory gene families of chordate genomes and focus on genes that were either not present or regarded as ‘orphan’ genes in the *Ciona* genome.

Materials and methods

Regulatory genes in the genome of *Branchiostoma floridae* (Putnam et al. 2008) were identified with the TBLASTN program (Altschul et al. 1990) using the *Ciona* genes and the human genes as queries. The best-fit gene models that were decided on the basis of homology and expressed sequence tag evidences were employed if several candidate gene models were predicted for a single locus. If no gene models were available, the corresponding genome sequences were manually translated. These models were subjected to reciprocal BLAST analyses against the human and *Ciona* proteomes (threshold *E* value=1E–05). Those that displayed reciprocal best-hit relationships both between the amphioxus and human proteins and between the amphioxus and *Ciona* proteins were considered most likely orthologs of these human and *Ciona* proteins, as the phylogenetic relationships between regulatory proteins of human and *Ciona* have been extensively examined (Hino et al. 2003; Satou et al. 2003a, b; Wada et al. 2003; Yagi et al. 2003; Yamada et al. 2003). The proteome sets are IPI ver3.24 for human and JGI ver1.0 for *Ciona*. Several subfamilies consist of the single members, and members of each subfamily contain a unique domain(s). In these cases, their orthologies are obvious by alignments, and therefore, we did not perform further molecular phylogenetic analyses. Id/Emc (group-D bHLH), EBF/COE (group-F bHLH), and some homeobox subfamilies are included in this category (ESM Fig. S6). Six orphan homeodomain proteins (70306/102512, 86186/124544, 99353, 102805, 106421, and 107468/124758) were highly diverged, and comparable sequences were not found in any known genomes. No gene models for a putative *Branchiostoma* ortholog of HMG2L included an HMG-box probably because of a sequence gap, and therefore, this protein was not included in the molecular phylogenetic tree. However, overall high similarities of this *Branchiostoma* protein with *Ciona* and human HMG2L protein suggested that this *Branchiostoma* protein is a bona fide ortholog of HMG2L, and therefore, it is tentatively included in ESM Table S1. For all of the other protein subfamilies, molecular phylogenetic analyses were performed. In molecular phylogenetic analysis, fly proteins was basically included for comparison, and, if necessary, proteins of other animals were included. Especially for

genes whose clear counterparts were not found in the representative animal genomes, we sampled as many proteins as possible from the public protein database and we considered them as candidate orthologs. The T-Coffee program was employed for generating alignments (Notredame et al. 2000). The alignments of conserved functional domains (e.g., homeodomains for homeobox proteins and bHLH domains for bHLH proteins) were used for constructing phylogenetic trees. Maximum likelihood trees were constructed with the PHYML program (Guindon and Gascuel 2003) with the WAG amino acid substitution matrix (Whelan and Goldman 2001), the proportion of invariant sites calculated from the alignment, and four rate categories with a gamma distribution parameter estimated from the data. Trees were tested with 100 bootstrap pseudoreplicates. Bayesian trees were constructed by the MRBAYES program (ver 3.1.2; Ronquist and Huelsenbeck 2003). Two independent runs were conducted (each with four chains) until the average standard deviation of split frequencies becomes less than 0.01 or until the two runs converge onto the stationary distribution. In cases that no clear orthologs of mammalian genes were found in the *Ciona* or *Branchiostoma* genome, we similarly examined whether their orthologous genes are found in the protozoans. In cases that no clear orthologs of *Ciona* and *Branchiostoma* genes were found in the human genome, we also similarly examined whether their orthologous genes are found in the genomes of mouse, *Xenopus tropicalis* and zebrafish. Molecular phylogenetic trees constructed in the present study are shown in ESM Figs. S1–S5. Multiple *Branchiostoma* gene models often encode very similar polypeptides. In most cases, it was hard to distinguish whether each of these genes represents haplotypes or whether these genes are due to assembly artifacts or are actually encoded in different loci (Putnam et al. 2008). Therefore, the number of the gene models shown in Table 2 is not equal to the total gene number of the regulatory genes.

Results and discussion

Regulatory genes encoding transcription factors and signaling molecules in major families were comprehensively surveyed on the amphioxus genome. The genes identified were classified based on results of molecular phylogenetic analyses (ESM Table S1). Comparisons of major transcription factor families and signaling ligands/receptors encoded in the *Ciona* and amphioxus genomes are summarized in Tables 1 and 2.

Only 26 gene families of the 311 examined were putatively vertebrate-specific (or partly mammalian or amniote-specific), as both of the amphioxus and *Ciona*

Table 1 Numbers of subfamilies of transcription factor genes encoded in the genomes of human, *Ciona*, and *Branchiostoma*

Human	<i>Ciona</i>	<i>Branchiostoma</i>	bHLH	bZIP	ETS	Fox	HMG	HMG/Sox	Homeobox	NR	T-box
+	+	+	33	18	10	19	10	8	73	16	7
+	+	–	3 (3)	1 (0)	0	0	0	0	1 (1)	1 (1)	0
+	–	+	9 (~6)	1 (1)	1 (1)	0	0	1 (0)	20 (17)	2 (2)	1 (0)
+	–	–	2 (1)	2 (0)	0	3 (0)	1 ^b (0)	3 (1)	16 (0)	1 (0)	0
–	+	+	1 (0)	0	0	1 (1 ^a)	0	0	2 (1)	0	0
–	+	–	~4 (0)	~4 (0)	0	~3 (0)	~2 (0)	0	3 (0)	0	0
–	–	+	0	~4 (0)	0	1 (0)	0	1 (0)	25 ^c (7)	2 (1)	0

Subfamilies are defined in ESM Table S1. Numbers of subfamilies found in at least one protostome genome were shown in parentheses.

^a Orthologs of *Ciona* Orphan Fox5 were found in *Fugu* in addition to *Drosophila*.

^b No counterparts belonging to this subfamily were found in non-mammalian genomes.

^c One of the genes is Hox14, whose clear ortholog is present in coelacanth and horn shark.

genomes did not have the counterparts. These putative vertebrate-specific genes include *GDF2/BMP10*, *DDIT3* (bZIP transcription factor), *SRY* (*SoxA*), and *FGF19* (Table 2). While most of these regulatory genes are represented in amphibians and fish, some genes, such as *SOHLH1/SOHLH2* (bHLH transcription factor) and *Fkhl18* (Fox transcription factor), lack an apparent ortholog in the genomes of amphibians and fish. These ‘missing’ genes are candidates for regulatory elements that emerged in the vertebrate evolutionary process.

There are a few genes found in the human and *Ciona* genomes but not in the amphioxus genome. These include *ACSC3*, *MESPI*, *MNT*, *BACH1*, *VDR*, *EFNA1*, and genes in three FGF subfamilies. However, because there are orthologs for *ACSC3*, *MESPI*, *MNT*, and *VDR* in protostomes, these genes appear to have been lost in the

amphioxus lineage. In contrast, other genes, such as *BACH1* (bZIP transcription factor) and *EFNA* (ephrin-A), are candidates that the common ancestors of *Ciona* and vertebrates acquired after the divergence of amphioxus and tunicates/vertebrates. In addition to *BACH1* and *EFNA*, we could not find any clear orthologs for three Fgf subfamilies in the amphioxus lineage. However, there are at least four ‘orphan’ Fgf genes in the *Branchiostoma* genome. It is possible that some of them belong to these subfamilies, but such orthologous relationships are hidden because of their rapid evolution in the amphioxus lineage. Of course, it is also possible that these Fgfs are ‘orphans’ and the amphioxus genome lacks orthologs for these FGFs.

Two types of ephrins are present in vertebrates; Ephrin-A proteins are GPI-anchored to the cell membrane and Ephrin-B proteins are transmembrane proteins (Kullander

Table 2 Regulatory genes potentially acquired or lost during the evolution of chordates

	Regulatory genes
Regulatory genes present in the human genome but not in <i>Ciona</i> , <i>Branchiostoma</i> , and <i>Drosophila</i> genomes ^a	<i>SOHLH1/SOHLH2</i> , <i>BATF/BATF2/SNFT</i> , <i>DDIT3</i> , <i>FKHL18</i> , <i>FOXL1</i> , <i>FOXR1/FOXR2</i> , <i>IPI00051152/IPI00472320/IPI00736639/IPI00739825</i> , <i>SOX15</i> , <i>SRY</i> , <i>ADNP/ADNP2</i> , <i>ARGFX</i> , <i>DPRX</i> , <i>ESX1</i> , <i>HESX1</i> , <i>LEUTX</i> , <i>LOC647589</i> , <i>MIXL</i> , <i>NANOG</i> , <i>NOBOX</i> , <i>POU5F1/POU5F2</i> , <i>RHOXF1/RHOXF2/RHOXF2B</i> , <i>SATB1/SATB2</i> , <i>SEBOX</i> , <i>TPRX1</i> , <i>VENTX</i> , <i>AR/NR3C1/NR3C2/PGR</i> , <i>GDF2/BMP10</i> , <i>FGF19/FGF21/FGF23</i>
Regulatory genes present in the human and the <i>Ciona</i> genomes but not in the amphioxus and protostome genomes ^a	<i>BACH1/BACH2</i> , <i>EFNA1/EFNA2/EFNA3/EFNA4/EFNA5</i> , <i>FGF11/FGF12/FGF13/FGF14</i> , <i>FGF3/FGF7/FGF10/FGF22</i> , <i>FGF4/FGF5/FGF6</i>
Regulatory genes present in the <i>Ciona</i> and amphioxus genomes but not in the human and protostome genomes ^a	<i>Nkx-A</i> , <i>ADMP</i> ^b , <i>Twist-like</i> ^c
Regulatory genes present in the genomes of <i>Ciona</i> , amphioxus, and protostomes, but not in the human genome ^a	<i>Orphan-Fox5</i> , <i>Nkx-C</i>
Regulatory genes present in the human and protostome genomes but not in the <i>Ciona</i> and amphioxus genomes ^a	<i>BBX</i> , <i>Twist</i> ^c

Genes belonging to the same subfamily are delimited by slashes.

^a Note that some genes not found in the present study are possibly ones that are encoded in the unsequenced genomic region or ones that have been rapidly evolved (thereby their orthologous relationships are hidden).

^b *ADMP* is also found in fish and amphibian genomes. See the text for detail.

^c *Ciona* Twist-like has a function similar to Twist, but orthology between Twist-like of *Ciona* and *Branchiostoma* and Twist of other animals are not necessarily supported by molecular phylogenetic methods.

and Klein 2002). In protostomes, the genome of *Drosophila melanogaster* does not contain any GPI-anchored ephrins, while the genome of *Caenorhabditis elegans* includes four potential GPI-anchored ephrins (Kullander and Klein 2002). However, molecular phylogenetic studies indicate that these protostome ephrins are not closely related to vertebrate ephrin-A or ephrin-B (Satou et al. 2003b). Thus, vertebrate ephrin-A and ephrin-B emerged after the split of protostomes and deuterostomes. The *Ciona* genome contains clear orthologs both of ephrin-A and ephrin-B (Satou et al. 2003b), while the amphioxus genome contains clear orthologs of ephrin-B, but not for ephrin-A. Therefore, it is likely that the last common ancestor of *Ciona* and vertebrates included an ephrin-A gene, but the common ancestor of all chordates did not. There are two types of genes encoding ephrin receptors (Eph), each of which binds to ephrin-A or ephrin-B. Molecular phylogenetic trees suggest that the amphioxus Ephs are more similar to those of nematode, insects, and sea urchins. This observation is consistent with lack of clear orthologs of ephrin-A in the amphioxus genome. However, it is possible that these genes were not included in the current assembly of the amphioxus genome or that these genes were lost in the amphioxus lineage.

There are a few genes that are found in the amphioxus and protostome genomes but not in the vertebrate and tunicate genomes. These include the homeobox genes *rough*, *reversed polarity*, *NK7.1*, *pox-neuro*, *defective proventriculus*, and a nuclear receptor, *NR5B*. The nuclear receptor family NR5 comprises two subfamilies, NR5A and NR5B. Genes belonging to the NR5A subfamily (*NR5A1* and *NR5A2* in human) are present in genomes from protostomes to vertebrates as well as in the *Ciona* and amphioxus genomes. In contrast, *NR5B* genes have only been observed in protostome genomes, while an ortholog of this gene is present in the amphioxus genome (Fig. 1). Therefore, it is likely that the ancestral chordate genome included this gene but it was lost after the split of amphioxus and tunicates/vertebrates in the tunicate/vertebrate lineage. We could not identify gene families that were present in the *Ciona* and protostome genomes but not in the vertebrate and amphioxus genomes.

An ortholog of an ‘orphan’ Fox gene present in the *Ciona* genome, *Ci-orphan-Fox5*, is present in the *Branchiostoma* genome. Molecular phylogenetic analyses revealed that this gene has a potential counterpart in insect (*D. melanogaster*) and fish (Fugu and Medaka) genomes but not in higher vertebrates (ESM Figs. S1 and S2). Although the function of this gene is unknown, its distribution may provide a clue for understanding vertebrate evolution. Similarly, *Nkx-C* was first identified in the *Ciona* genome and is present in the fly and amphioxus genome but not in vertebrate genomes. Therefore, these genes

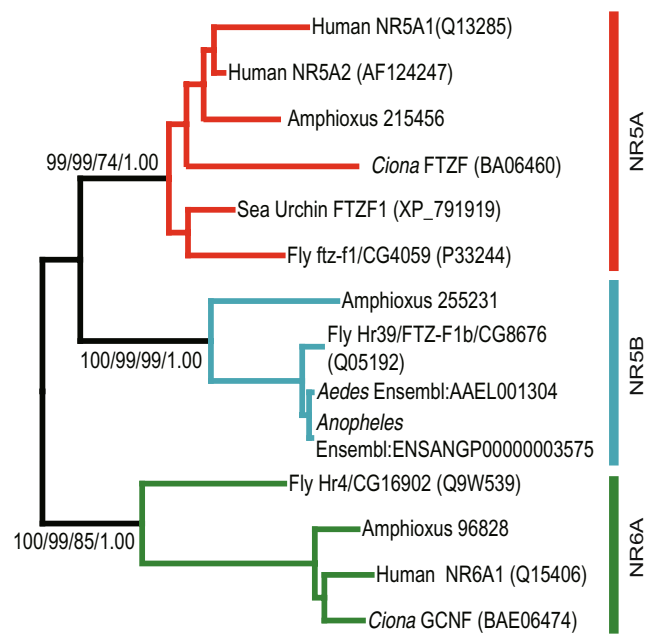


Fig. 1 A molecular phylogenetic tree of nuclear receptors (NR5 and NR6 subfamilies) constructed by the maximum likelihood method. Four numbers of each of three major nodes indicate bootstrap values by the maximum likelihood, neighbor-joining, maximum parsimony methods, and posterior probabilities by the Bayesian method. All of four methods confirm that the amphioxus genome has a gene encoding a protein belonging to the NR5B subfamily

emerged before the split of deuterostomes and protostomes, but were lost in the vertebrate lineage. In comparison, three genes, namely *Nkx-A*, *ADMP*, and *Twist-like* are not found in protostome genomes. While *ADMP* is present in *fugu* and *Xenopus* genomes, *Nkx-A* and *Twist-like* are found only in the *Ciona* and amphioxus genomes. The *Twist-like* genes in these species are orthologs, but this relationship does not extend to Twist proteins of vertebrates and protostomes (ESM Fig. S5), although *Ciona* Twist-like has a similar function to Twist (Imai et al. 2003) and both of the *Ciona* and *Branchiostoma* genomes lack a clear Twist ortholog.

In spite of extensive molecular phylogenetic analyses, orphan regulatory genes remain both in the tunicate genome and in the amphioxus genome. Some of these genes may have been rapidly evolved in these species and thus obscured their similarity to regulatory genes in other organisms. Alternatively, the ancestral chordate may have possessed these orphan genes, but they were lost during subsequent evolution. In some cases, syntenic relationships provide strong evidence for orthologous relationships that are otherwise impossible to resolve by the molecular phylogeny. For example, the orphan TGF β gene, called as *TGF β -NA1(OrphanTGF β -1)*, is encoded in the first intron of another TGF β gene (*Ci-BMP2/4*, an ortholog of vertebrate BMP2 and BMP4) in the *Ciona* genome (Fig. 2). This gene organization may be ancestral situation

within deuterostomes. Previous neighbor-joining molecular phylogenetic analysis failed to assign *Ciona* TGF β -NA1 to a specific subfamily (Hino et al. 2003), and the present maximum likelihood and Bayesian trees weakly indicated that this orphan TGF β is grouped with sea urchin Univin, zebrafish DVR-1, frog VG1, and human GDF1 and GDF3 (ESM Figs. S3 and S4). In the present survey on the amphioxus genome, we discovered a TGF β gene (tentatively named *GDF1/3-like2*) encoded in the first intron of *BMP2/4* gene in this species (Fig. 2). Similar genetic organization is present in the sea urchin and zebrafish genomes. *Univin*, a BMP2/4 related TGF β , is found to be a 5'-neighbor of *BMP2/4* in the sea urchin genome (Fig. 2; Sodergren et al. 2006). *DVR-1* is a 5'-neighbor of *BMP2a* in the zebrafish genome (Fig. 2). However, it is possible that there are additional exons of BMP2 that have not yet been identified because of insufficient experimental data, and

univin and *DVR-1* are also encoded in the first intron of *BMP2/4*. Additionally, the gene 5' to the amphioxus *GDF1/3-like2/BMP2/4* is an ortholog of human *FBXW7*. The *Ciona* ortholog of *FBXW7* is also the 5' neighbor of these two TGF β genes. The gene 3' to these TGF β genes in the amphioxus genome encodes a receptor tyrosine kinase. The best candidate in the sea urchin genome for the ortholog of this amphioxus RTK gene was 5' neighbor of *univin*. This genetic organization supports the conclusion that these genes, sea urchin *univin*, *Ciona* TGF β -NA1, *Branchiostoma* *GDF1/3-like2*, and zebrafish *DVR-1*, are orthologous for each other.

This genetic organization also implies that the founder gene of the subfamily including *univin*, *Ciona* TGF β -NA1, *Branchiostoma* *GDF1/3-like2*, and zebrafish *DVR-1* arose by gene duplication of *BMP2/4* at the emergence of deuterostome. This hypothesis is supported by the molecular

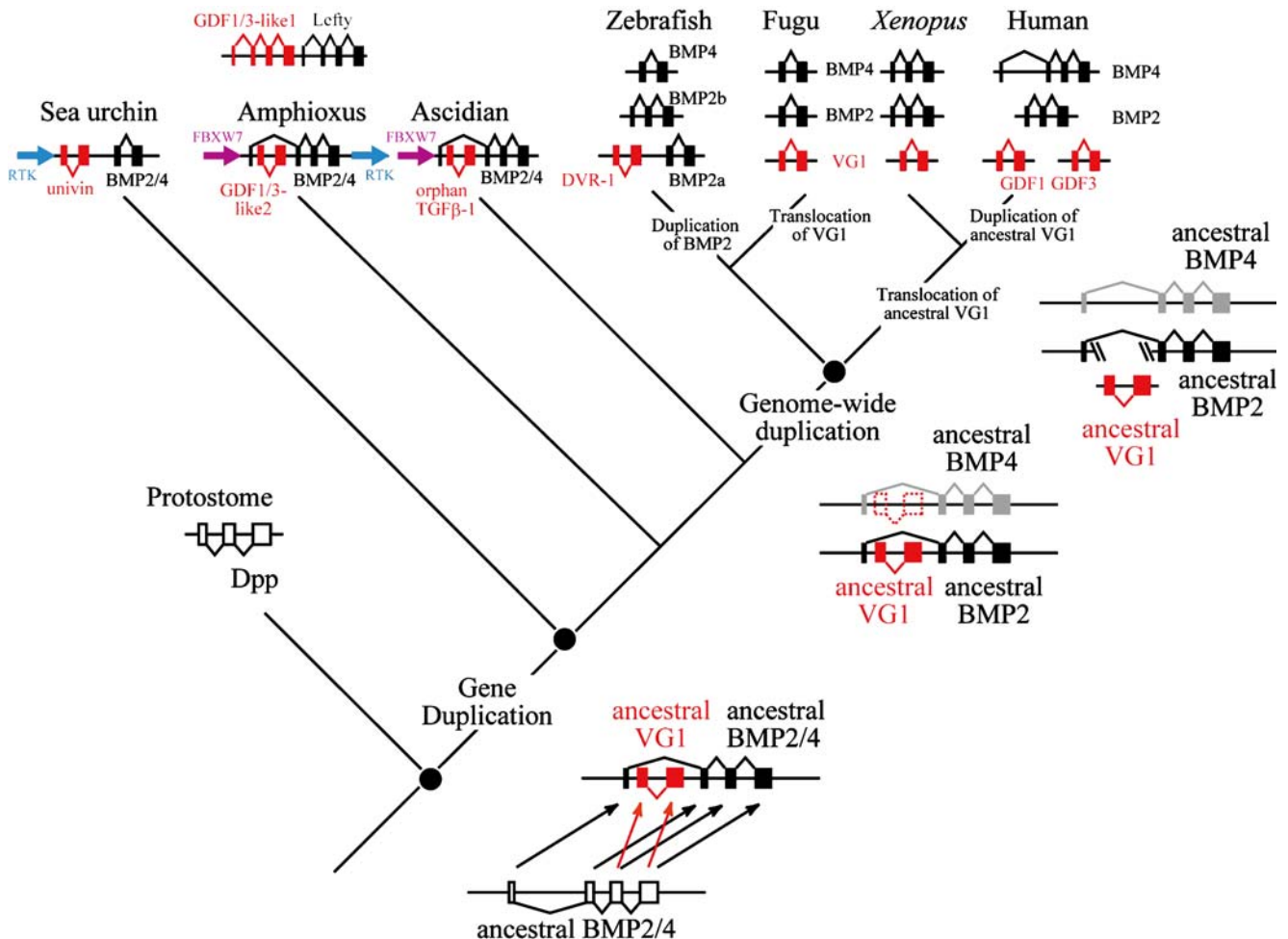


Fig. 2 A possible origin of *GDF1/GDF3*, or *VG1*, and *Lefty*. *GDF1/GDF3* and the predicted orthologs are depicted in red. *BMP2/BMP4* and its orthologs are depicted in black. Cyan and blue arrows

represent genes encoding an F-box/WD-repeated protein and a receptor tyrosine kinase, respectively. The synteny of these genes was partially conserved in the tunicate and sea urchin genome

phylogenetic trees that imply a close relation between univin/TGF β -NA1/GDF1/3-like2/DVR-1 and BMP2/BMP4 (ESM Figs. S3 and S4). Zebrafish *DVR-1* is an ortholog of frog *VG1/Derrier* and human *GDF1/GDF3*, which are not located near *BMP2* or *BMP4* in these species (Fig. 2). The founder gene arisen by gene duplication of *BMP2/4* might be translocated in the most vertebrate genomes.

An additional ‘orphan’ TGF β was also identified in the amphioxus genome (tentatively named as *GDF1/3-like1*). This gene product is most similar to *GDF1/3-like2* and secondarily to *BMP2/4* (ESM Figs. S3 and S4). This gene is located next to *Lefty*, a divergent member of the TGF β superfamily encoding an antagonist for another TGF β , *Nodal* (Fig. 2). Because *Lefty* and *Nodal* are found only in the deuterostome genomes, these genes likely evolved from the other TGF β at the emergence of the ancestral deuterostome. Previous studies demonstrate that acquisition and evolution of the Nodal system is one of the key events of deuterostome evolution (Duboc et al. 2004). Although molecular phylogenetic analyses failed to identify the origins of *Lefty*, it is very unlikely that these two related genes, *GDF1/3-like1* and *Lefty*, are located next to each other by chance. It is also unlikely that some evolutionary constraint affects this gene pair exclusively in this species. Therefore, we propose that *Lefty* arose by two duplications of the ancestral *GDF1/3* or by a duplication of *BMP2/4*-*GDF1/3* gene pairs.

The counterparts for most of orphan *Ciona* genes are not found in the amphioxus genome (ESM Table S1). For example, there is a gene called *Notrlc* in the *Ciona* genome. This gene is similar to *Hand*, but a bona fide *Hand* ortholog is encoded at a distinct locus. The absence of the ortholog of the *Ciona* orphan genes in the *Branchiostoma* genome suggests that most of these genes evolved or arose after the split of the tunicate and vertebrate lineages, although it cannot be denied that the ancestral chordate had some part of these genes and *Branchiostoma* and vertebrates lost them independently.

Many regulatory genes not present in the *Ciona* genome were identified in the amphioxus genome. On the other hand, 46 genes in 26 regulatory gene families were not discovered in the *Ciona* and *Branchiostoma* genomes, although it is still possible that a fraction of these ‘missing’ genes are included in unsequenced genomic regions. In summary, the ancestral chordate appears to have had a repertory of regulatory genes similar to that of modern vertebrates, but it is possible that approximately 10% of regulatory genes in the vertebrate genome were generated after the emergence of vertebrates. Together with the emergence of these regulatory genes, the emergence of paralogous genes for each subfamily in vertebrate genome (2.3 on average) have undoubtedly made important contributions to the evolution of modern vertebrates. As the

Ciona genome has lost many regulatory genes, the regulatory gene set for the developmental program required to form tadpole type larva may be smaller than the gene set found in *Ciona* and *Branchiostoma* genomes.

Acknowledgments The present study was supported by the Grant for the Biodiversity Research of the 21st Century COE (A14) from MEXT, Japan.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Dehal P, Satou Y, Campbell RK et al (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298:2157–2167
- Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439:965–968
- Duboc V, Rottinger E, Besnardeau L, Lepage T (2004) Nodal and BMP2/4 signaling organizes the oral–aboral axis of the sea urchin embryo. *Dev Cell* 6:397–410
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704
- Hino K, Satou Y, Yagi K, Satoh N (2003) A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. VI. Genes for Wnt, TGF β , Hedgehog and JAK/STAT signaling pathways. *Dev Genes Evol* 213:264–272
- Imai KS, Satoh N, Satou Y (2003) A Twist-like bHLH gene is a downstream factor of an endogenous FGF and determines mesenchymal fate in the ascidian embryos. *Development* 130:4461–4472
- Kullander K, Klein R (2002) Mechanisms and functions of Eph and ephrin signalling. *Nat Rev Mol Cell Biol* 3:475–486
- Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217
- Putnam NH, Blow MJ, Butts T et al (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* (in press)
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
- Satou Y, Imai KS, Levine M, Kohara Y, Rokhsar D, Satoh N (2003a) A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. I. Genes for bHLH transcription factors. *Dev Genes Evol* 213:213–221
- Satou Y, Sasakura Y, Yamada L, Imai KS, Satoh N, Degnan B (2003b) A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. V. Genes for receptor tyrosine kinase pathway and Notch signaling pathway. *Dev Genes Evol* 213:254–263
- Sodergren E, Weinstock GM, Davidson EH et al (2006) The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314:941–952
- Wada S, Tokuoka M, Shoguchi E, Kobayashi K, Di Gregorio A, Spagnuolo A, Branno M, Kohara Y, Rokhsar D, Levine M, Saiga H, Satoh N, Satou Y (2003) A genomewide survey of

- developmentally relevant genes in *Ciona intestinalis*. II. Genes for homeobox transcription factors. Dev Genes Evol 213:222–234
- Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699
- Yagi K, Satou Y, Mazet F, Shimeld SM, Degnan B, Rokhsar D, Levine M, Kohara Y, Satoh N (2003) A genome-wide survey of developmentally relevant genes in *Ciona intestinalis*. III. Genes for Fox, ETS, nuclear receptors and NFkappaB. Dev Genes Evol 213:235–244
- Yamada L, Kobayashi K, Degnan B, Satoh N, Satou Y (2003) A genome-wide survey of developmentally relevant genes in *Ciona intestinalis*. IV. Genes for HMG transcriptional regulators, bZip and GATA/Gli/Zic/Snail. Dev Genes Evol 213:245–253